

Strong Consistency of the Sign-Perturbed Sums Method

Balázs Csanád Csáji*

Marco C. Campi†

Erik Weyer‡

Abstract—Sign-Perturbed Sums (SPS) is a recently developed non-asymptotic system identification algorithm that constructs confidence regions for parameters of dynamical systems. It works under mild statistical assumptions, such as symmetric and independent noise terms. The SPS confidence region includes the least-squares estimate, and, for any finite sample and user-chosen confidence probability, the constructed region contains the true system parameter with exactly the given probability. The main contribution in this paper is to prove that SPS is strongly consistent, in case of linear regression based models, in the sense that any false parameter will almost surely be excluded from the confidence region as the sample size tends to infinity. The asymptotic behavior of the confidence regions constructed by SPS is also illustrated by numerical experiments.

I. INTRODUCTION

Mathematical models of dynamical systems are of widespread use in many fields of science, engineering and economics. Such models are often obtained using system identification techniques, that is, the models are estimated from observed data. There will always be uncertainty associated with models of dynamical systems, and an important problem is the uncertainty evaluation of models.

Previously, the Sign-Perturbed Sums (SPS) algorithm was introduced for linear systems [1], [2], [3], [4], [8]. The main feature of the SPS method is that it constructs a confidence region which has an exact probability of containing the true system parameter based on a finite number of observed data. Moreover, the least-squares estimate of the true parameter belongs to the confidence region. In contrast with asymptotic theory of system identification, e.g., [5], which only delivers confidence ellipsoids that are guaranteed asymptotically as the number of data points tends to infinity, the SPS regions are guaranteed for any finite number of data points.

Although the main draw card of SPS is the finite sample properties, the asymptotic properties are also of interests. One of the fundamental asymptotic properties a confidence region construction can have is consistency [6], which indicate that false parameter values will eventually be “filtered

The work of B. Cs. Csáji was partially supported by the Australian Research Council (ARC) under the Discovery Early Career Researcher Award (DECRA) DE120102601 and by the János Bolyai Research Fellowship of the Hungarian Academy of Sciences, contract no. BO/00683/12/6. The work of M. C. Campi was partly supported by MIUR - Ministero dell’Istruzione, dell’Università e della Ricerca. The work of E. Weyer was supported by the ARC under Discovery Grants DP0986162 and DP130104028.

*MTA SZTAKI: Institute for Computer Science and Control, Hungarian Academy of Sciences, Kende utca 13–17, Budapest, Hungary, H-1111; balazs.csaji@sztaki.mta.hu

†Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy; campi@ing.unibs.it

‡Department of Electrical and Electronic Engineering, Melbourne School of Engineering, The University of Melbourne, 240 Grattan Street, Parkville, Melbourne, Victoria, 3010, Australia; ewey@unimelb.edu.au

out” as we have more and more data. In this paper we show that SPS is in fact *strongly consistent*, i.e., the SPS confidence region shrinks around the true parameter as the sample size increases and, asymptotically, any false parameter will almost surely be excluded from the confidence region.

Besides the theoretical analysis, we also include a simulation example which illustrates the behavior of the SPS confidence region as the number of data points increases.

The paper is organized as follows. In the next section we briefly summarize the problem setting, our main assumptions, the SPS algorithm and its ellipsoidal outer-approximation. The strong consistency results are given in Section III, and they are illustrated on a simulation example in Section IV. The proofs can be found in the appendices.

II. THE SIGN-PERTURBED SUMS METHOD

We start by briefly summarizing the SPS method for linear regression problems. For more details, see [2], [3], [8].

A. Problem Setting

The data is generated by the following system

$$Y_t \triangleq \varphi_t^T \theta^* + N_t,$$

where Y_t is the output, N_t is the noise, φ_t is the regressor, θ^* is the unknown true parameter and t is the time index. Y_t and N_t are scalars, while φ_t and θ^* are d dimensional vectors. We consider a sample of size n which consists of the regressors $\varphi_1, \dots, \varphi_n$ and the outputs Y_1, \dots, Y_n . We aim at building a guaranteed confidence region for θ^* .

B. Main Assumptions

The assumptions on the noise and the regressors are

- A1 $\{N_t\}$ is a sequence of independent random variables. Each N_t has a symmetric probability distribution about zero, i.e., N_t and $-N_t$ has the same distribution.

- A2 Each regressor, φ_t , is deterministic and

$$R_n \triangleq \frac{1}{n} \sum_{t=1}^n \varphi_t \varphi_t^T.$$

is non-singular.

Note the weak assumptions, e.g., the noise terms can be nonstationary with unknown distributions and there are no moment or density requirements either. The symmetry assumption is also mild, as many standard distributions, including Gaussian, Laplace, Cauchy-Lorentz, Bernoulli, Binomial, Students t, logistic and uniform satisfy this property.

The restriction on the regressor vectors allow dynamical systems, for example, with transfer functions

$$G(z, \theta^*) = \sum_{k=1}^d \theta_k^* L_k(z, \beta),$$

where z is the shift operator and $\{L_k(z, \beta)\}$ is a function expansion with a (fixed) user-chosen parameter β . The regressors in this case are $\varphi_t = [L_1(z, \beta) u_t, \dots, L_d(z, \beta) u_t]$, where $\{u_t\}$ is an input signal. Using $L_k(z, \beta) = z^{-k}$ corresponds to the standard FIR model, while more sophisticated choices include Laguerre-, and Kautz basis functions [5], [7], which are often used to model (or approximate) systems with slowly decaying impulse responses

C. Intuitive Idea of SPS

We note that the least-squares estimate of θ^* is given by

$$\hat{\theta}_n \triangleq \arg \min_{\theta \in \mathbb{R}^d} \sum_{t=1}^n (Y_t - \varphi_t^T \theta)^2.$$

which can be found by solving the *normal equation*, i.e.,

$$\sum_{t=1}^n \varphi_t (Y_t - \varphi_t^T \theta) = 0,$$

The main building block of the SPS algorithm is, as its name suggests, $m-1$ sign-perturbed versions of the normal equation (which are also normalized by $\frac{1}{n} R_n^{-1/2}$). More precisely, the sign-perturbed sums are defined as follows

$$S_i(\theta) \triangleq R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t \alpha_{i,t} (Y_t - \varphi_t^T \theta),$$

$i \in \{1, \dots, m-1\}$, and a reference sum is given by

$$S_0(\theta) \triangleq R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t (Y_t - \varphi_t^T \theta).$$

Here $R_n^{-\frac{1}{2}}$ is such that $R_n = R_n^{-\frac{1}{2}} R_n^{-\frac{1}{2}T}$, and $\alpha \triangleq \{\alpha_{i,t}\}$ are independent and identically distributed (i.i.d.) Rademacher variables, i.e., they take ± 1 with probability $1/2$ each.

A key observation is that for $\theta = \theta^*$

$$S_0(\theta^*) = R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t N_t,$$

$$S_i(\theta^*) = R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t N_t = R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \pm \varphi_t N_t.$$

As $\{N_t\}$ are independent and symmetric, there is no reason why $\|S_0(\theta^*)\|^2$ should be bigger or smaller than any another $\|S_i(\theta^*)\|^2$ and this is utilized by SPS by excluding those values of θ for which $\|S_0(\theta)\|^2$ is among the q largest ones, and as stated below, the so constructed confidence set has exact probability $1 - q/m$ of containing the true parameter. It can also be noted that when $\theta - \theta^*$ is large, $\|S_0(\theta)\|^2$ tends to be the largest one of the m functions, such that values far away from θ^* are excluded from the confidence set.

PSEUDOCODE: SPS-INITIALIZATION

1. Given a confidence probability $p \in (0, 1)$, set integers $m > q > 0$ such that $p = 1 - q/m$;
2. Calculate the

$$R_n \triangleq \frac{1}{n} \sum_{t=1}^n \varphi_t \varphi_t^T;$$
 and find a factor $R_n^{1/2}$ such that

$$R_n^{1/2} R_n^{1/2T} = R_n;$$
3. Generate $n(m-1)$ i.i.d. random signs $\{\alpha_{i,t}\}$ with

$$\mathbb{P}(\alpha_{i,t} = 1) = \mathbb{P}(\alpha_{i,t} = -1) = \frac{1}{2},$$
 for $i \in \{1, \dots, m-1\}$ and $t \in \{1, \dots, n\}$;
4. Generate a random permutation π of the set $\{0, \dots, m-1\}$, where each of the $m!$ permutations has the same probability $1/(m!)$ to be selected.

TABLE I

PSEUDOCODE: $\mathbb{I}_{\text{SPS}}(\theta) \sim \text{SPS-INDICATOR}(\theta)$

1. For the given θ evaluate

$$S_0(\theta) \triangleq R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t (Y_t - \varphi_t^T \theta),$$

$$S_i(\theta) \triangleq R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t (Y_t - \varphi_t^T \theta),$$
 for $i \in \{1, \dots, m-1\}$;
2. Order scalars $\{\|S_i(\theta)\|^2\}$ according to \succ_{π} ;
3. Compute the rank $\mathcal{R}(\theta)$ of $\|S_0(\theta)\|^2$ in the ordering, where $\mathcal{R}(\theta) = 1$ if $\|S_0(\theta)\|^2$ is the smallest in the ordering, $\mathcal{R}(\theta) = 2$ if $\|S_0(\theta)\|^2$ is the second smallest, and so on;
4. Return 1 if $\mathcal{R}(\theta) \leq m - q$, otherwise return 0.

TABLE II

D. Formal Construction of the SPS Confidence Region

The pseudocode of the SPS algorithm is presented in two parts. The initialization (Table I) sets the main global parameters and generates the random objects needed for the construction. In the initialization, the user provides the desired confidence probability p . The second part (Table II) evaluates an indicator function, $\mathbb{I}_{\text{SPS}}(\theta)$, which determines if a particular parameter θ belongs to the confidence region.

The permutation π generated in the initialization defines a strict total order \succ_{π} which is used to break ties in case two $\|S_i(\theta)\|^2$ functions take on the same value. Given m scalars Z_0, \dots, Z_{m-1} , relation \succ_{π} is defined by

$$Z_k \succ_{\pi} Z_j \quad \text{if and only if} \\ (Z_k > Z_j) \quad \text{or} \quad (Z_k = Z_j \quad \text{and} \quad \pi(k) > \pi(j)).$$

The p -level SPS confidence region is given by

$$\hat{\Theta}_n \triangleq \{ \theta \in \mathbb{R}^d : \mathbb{I}_{\text{SPS}}(\theta) = 1 \}.$$

Note that the least-squares estimate (LSE), $\hat{\theta}_n$, has the property that $S_0(\hat{\theta}_n) = 0$. Therefore, the LSE is included in the SPS confidence region, assuming that it is non-empty.

As was shown¹ in [2], the most important property of the SPS method is that the constructed confidence region contains θ^* with exact probability p , more precisely

Theorem 1: Assuming A1 and A2, the confidence probability of the constructed SPS region is exactly p , that is,

$$\mathbb{P}(\theta^* \in \hat{\Theta}_n) = 1 - \frac{q}{m} = p.$$

Since the confidence probability is *exact*, no conservatism is introduced, despite the mild statistical assumptions.

E. Ellipsoidal Outer-Approximation

Given a particular value of θ , it is easy to check whether θ is in the confidence region, i.e., we simply need to evaluate the indicator function at θ . Hence, SPS is well suited to problems where only a finite number of θ values need to be checked. This is, e.g., the case in some hypothesis testing and fault detection problems. On the other hand, it can be computationally demanding to construct the boundary of the region. E.g. evaluating the indicator function on a grid, suffers from the ‘‘curse of dimensionality’’. Now we briefly recall an approximation algorithm for SPS, suggested in [8], which can be efficiently computed and offers a compact representation in the form of ellipsoidal over-bounds.

After some manipulations [8] we can write $\|S_0(\theta)\|^2$ as

$$\|S_0(\theta)\|^2 = (\theta - \hat{\theta}_n)^T R_n (\theta - \hat{\theta}_n),$$

thus, the SPS region is given by those values of θ that satisfy

$$\hat{\Theta}_n = \left\{ \theta \in \mathbb{R}^d : (\theta - \hat{\theta}_n)^T R_n (\theta - \hat{\theta}_n) \leq r(\theta) \right\},$$

where $r(\theta)$ is the q th largest value of the functions $\|S_i(\theta)\|^2$, $i \in \{1, \dots, m-1\}$. The idea is now to seek an over-bound by replacing $r(\theta)$ with a θ independent r , i.e.,

$$\hat{\Theta}_n \subseteq \left\{ \theta \in \mathbb{R}^d : (\theta - \hat{\theta}_n)^T R_n (\theta - \hat{\theta}_n) \leq r \right\}.$$

This outer-approximation will have the same shape and orientation as the standard asymptotic confidence ellipsoid [5], but it will have a different volume.

F. Convex Programming Formulation

In [8] it was shown that such an ellipsoidal over-bound can be constructed (Table III) by solving $m-1$ convex optimization problems. More precisely, if we compare $\|S_0(\theta)\|^2$ with one single $\|S_i(\theta)\|^2$ function, we have

$$\begin{aligned} & \left\{ \theta : \|S_0(\theta)\|^2 \leq \|S_i(\theta)\|^2 \right\} \\ & \subseteq \left\{ \theta : \|S_0(\theta)\|^2 \leq \max_{\theta: \|S_0(\theta)\|^2 \leq \|S_i(\theta)\|^2} \|S_0(\theta)\|^2 \right\}. \end{aligned}$$

¹Theorem 1 was originally proved using a slightly different tie-breaking approach, however, this does not affect the confidence probability.

PSEUDOCODE: SPS-OUTER-APPROXIMATION

1. Compute the least-squares estimate,
$$\hat{\theta}_n = R_n^{-1} \left[\frac{1}{n} \sum_{t=1}^n \varphi_t Y_t \right];$$
2. For $i \in \{1, \dots, m-1\}$, solve the optimization problem (1), and let γ_i^* be the optimal value;
3. Let r_n be the q th largest γ_i^* value;
4. The outer approximation of the SPS confidence region is given by the ellipsoid
$$\hat{\hat{\Theta}}_n = \left\{ \theta \in \mathbb{R}^d : (\theta - \hat{\theta}_n)^T R_n (\theta - \hat{\theta}_n) \leq r_n \right\}.$$

TABLE III

The maximization on the right-hand side generally leads to a nonconvex problem, however, its dual is convex and strong duality holds [8]. Hence, it can be computed by

$$\begin{aligned} & \text{minimize} && \gamma \\ & \text{subject to} && \lambda \geq 0 \\ & && \begin{bmatrix} -I + \lambda A_i & \lambda b_i \\ \lambda b_i^T & \lambda c_i + \gamma \end{bmatrix} \succeq 0, \end{aligned} \quad (1)$$

where relation ‘‘ $\succeq 0$ ’’ denotes that a matrix is positive semidefinite and A_i , b_i and c_i are defined as follows

$$\begin{aligned} A_i & \triangleq I - R_n^{-\frac{1}{2}} Q_i R_n^{-1} Q_i R_n^{-\frac{1}{2}T}, \\ b_i & \triangleq R_n^{-\frac{1}{2}} Q_i R_n^{-1} (\psi_i - Q_i \hat{\theta}_n), \\ c_i & \triangleq -\psi_i^T R_n^{-1} \psi_i + 2\hat{\theta}_n^T Q_i R_n^{-1} \psi_i - \hat{\theta}_n^T Q_i R_n^{-1} Q_i \hat{\theta}_n, \\ Q_i & \triangleq \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t \varphi_t^T, \\ \psi_i & \triangleq \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t Y_t. \end{aligned}$$

Letting γ_i^* be the value of program (1), we now have

$$\left\{ \theta : \|S_0(\theta)\|^2 \leq \|S_i(\theta)\|^2 \right\} \subseteq \left\{ \theta : \|S_0(\theta)\|^2 \leq \gamma_i^* \right\}.$$

Consequently, an outer approximation can be constructed by

$$\hat{\hat{\Theta}}_n \subseteq \hat{\Theta}_n \triangleq \left\{ \theta \in \mathbb{R}^d : (\theta - \hat{\theta}_n)^T R_n (\theta - \hat{\theta}_n) \leq r_n \right\},$$

where $r_n = q$ th largest value of γ_i^* , $i \in \{1, \dots, m-1\}$.

$\hat{\hat{\Theta}}_n$ is an ellipsoidal over-bound and it is also clear that

$$\mathbb{P}(\theta^* \in \hat{\hat{\Theta}}_n) \geq 1 - \frac{q}{m} = p,$$

for any finite n . Hence, the confidence ellipsoids based on SPS are rigorously guaranteed for finite samples, even though the noise may be nonstationary with unknown distributions.

III. STRONG CONSISTENCY

In addition to the probability of containing the true parameter, another important aspect is the size of the confidence set. While for a finite sample this generally depends on the

characteristics of the noise, here we show that (asymptotically) the SPS algorithm is *strongly consistent* in the sense that its confidence regions shrink around the true parameter, as the sample size increases, and eventually exclude any other parameters $\theta' \neq \theta^*$ with probability one.

We will use the following additional assumptions:

A3 *There exists a positive definite matrix R such that*

$$\lim_{n \rightarrow \infty} R_n = R.$$

A4 *(regressor growth rate restriction)*

$$\sum_{t=1}^{\infty} \frac{\|\varphi_t\|^4}{t^2} < \infty.$$

A5 *(noise variance growth rate restriction)*

$$\sum_{t=1}^{\infty} \frac{(\mathbb{E}[N_t^2])^2}{t^2} < \infty.$$

In the theorem below, $B_\varepsilon(\theta^*)$ denotes the usual norm-ball centered at θ^* with radius $\varepsilon > 0$, i.e.,

$$B_\varepsilon(\theta^*) \triangleq \{\theta \in \mathbb{R}^d : \|\theta - \theta^*\| \leq \varepsilon\}.$$

Theorem 2 states that the confidence regions $\{\hat{\Theta}_n\}$ eventually (almost surely) will be included in any norm-ball centered at θ^* as the sample size increases.

Theorem 2: Assuming A1, A2, A3, A4 and A5 : $\forall \varepsilon > 0$, there exists (a.s.) an N , such that $\forall n > N : \hat{\Theta}_n \subseteq B_\varepsilon(\theta^)$.*

The proof of Theorem 2 can be found in Appendix I. $N = N(\omega)$, that is, the actual sample size for which the confidence regions will remain inside an ε norm-ball around the true parameter depends on the noise realization.

Note that also for this asymptotic result, the noise terms can be nonstationary and their variances can grow to infinity, as long as their growth-rate satisfy condition A5. Also, the magnitude of the regressors can grow without bound, as long as it does not grow too fast, as controlled by A4.

Based on the proof, we can also conclude that

Corollary 3: Under the assumptions of Theorem 2, the radii, $\{r_n\}$, of the ellipsoidal outer-approximations, $\{\hat{\Theta}_n\}$, almost surely converge to zero as $n \rightarrow \infty$.

The proof sketch of this claim is given in Appendix II. Note that we already know [5] that the centers of the ellipsoidal over-bounds, $\{\hat{\theta}_n\}$, i.e., the LSEs, converge (a.s.) to θ^* .

IV. SIMULATION EXAMPLE

In this section we illustrate with simulations the asymptotic behavior of SPS and its ellipsoidal over-bound.

A. Second Order FIR System

We consider the following second order FIR system

$$Y_t = b_1^* U_{t-1} + b_2^* U_{t-2} + N_t,$$

where $b_1^* = 0.7$ and $b_2^* = 0.3$ are the true system parameters and $\{N_t\}$ is a sequence of i.i.d. Laplacian random variables with zero mean and variance 0.1. The input signal is

$$U_t = 0.75 U_{t-1} + W_t,$$

where $\{W_t\}$ is a sequence i.i.d. Gaussian random variables with zero mean and variance 1. The predictors are given by

$$\hat{Y}_t(\theta) = b_1 U_{t-1} + b_2 U_{t-2} = \varphi_t^T \theta,$$

where $\theta = [b_1, b_2]^T$ is the model parameter (vector), and $\varphi_t = [U_{t-1}, U_{t-2}]^T$ is the regressor vector.

Initially we construct a 95% confidence region for $\theta^* = [b_1^*, b_2^*]^T$ based on $n = 25$ data points, namely, $(Y_t, \varphi_t) = (Y_t, [U_{t-1}, U_{t-2}]^T)$, $t \in \{1, \dots, 25\}$.

We compute the shaping matrix

$$R_{25} = \frac{1}{25} \sum_{t=1}^{25} \begin{bmatrix} U_{t-1} \\ U_{t-2} \end{bmatrix} [U_{t-1} \ U_{t-2}],$$

and find a factor $R_{25}^{\frac{1}{2}}$ such that $R_{25}^{\frac{1}{2}} R_{25}^{\frac{1}{2}T} = R_{25}$. Then, we compute the reference sum

$$S_0(\theta) = R_{25}^{-\frac{1}{2}} \frac{1}{25} \sum_{t=1}^{25} \begin{bmatrix} U_{t-1} \\ U_{t-2} \end{bmatrix} (Y_t - b_1 U_{t-1} - b_2 U_{t-2}),$$

and using $m = 100$ and $q = 5$, we compute the 99 sign perturbed sums, $i \in \{1, \dots, 99\}$

$$S_i(\theta) = R_{25}^{-\frac{1}{2}} \frac{1}{25} \sum_{t=1}^{25} \begin{bmatrix} U_{t-1} \\ U_{t-2} \end{bmatrix} \alpha_{i,t} (Y_t - b_1 U_{t-1} - b_2 U_{t-2}),$$

where $\alpha_{i,t}$ are i.i.d. random signs. The confidence region is constructed as the values of θ for which at least $q = 5$ of the $\{\|S_i(\theta)\|^2\}$, $i \neq 0$, functions are larger (w.r.t. \succ_π) than $\|S_0(\theta)\|^2$. It follows from Theorem 1 that the confidence region constructed by SPS contains the true parameter with exact probability $1 - \frac{5}{100} = 0.95$.

The SPS confidence region is shown in Figure 1 together with the approximate confidence ellipsoid based on asymptotic system identification theory (with the noise variance estimated as $\hat{\sigma}^2 = \frac{1}{23} \sum_{t=1}^{25} (Y_t - \varphi_t^T \hat{\theta}_n)^2$).

It can be observed that the non-asymptotic SPS regions are similar in size and shape to the asymptotic confidence regions, but have the advantage that they are guaranteed to contain the true parameter with exact probability 0.95.

Next, the number of data points were increased to $n = 400$, still with $q = 5$ and $m = 100$, and the confidence regions in Figure 2 were obtained. As can be seen, the SPS confidence region concentrates around the true parameter as n increases. This is further illustrated in Figure 3 where the number of data points has been increase to 6400. Now, there is very little difference between the SPS confidence region, its outer approximation and the confidence ellipsoid based on asymptotic theory demonstrating the convergence result.

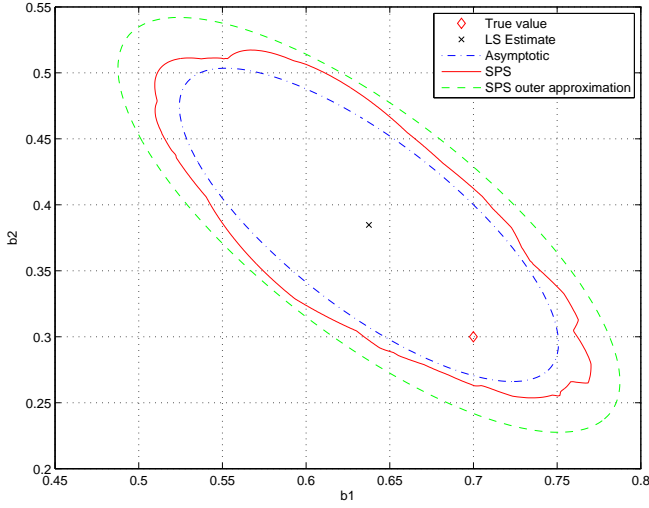


Fig. 1. 95% confidence regions, $n = 25$, $m = 100$, $q = 5$.

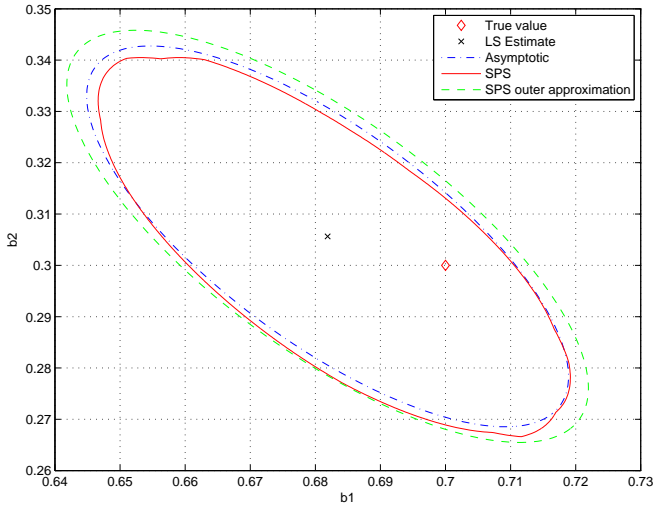


Fig. 2. 95% confidence regions, $n = 400$, $m = 100$, $q = 5$.

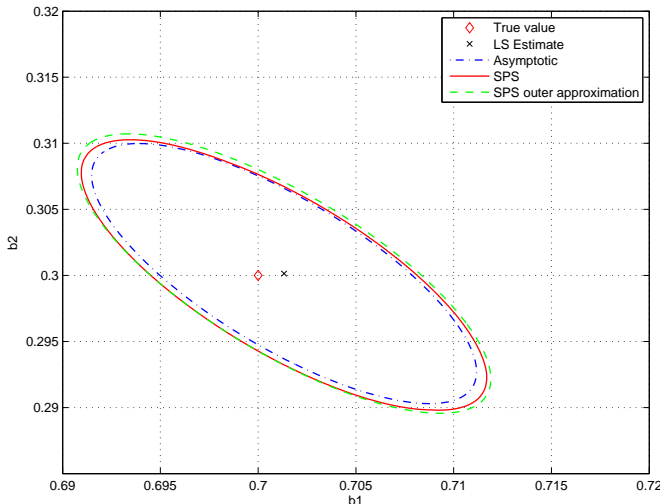


Fig. 3. 95% confidence regions, $n = 6400$, $m = 100$, $q = 5$.

V. SUMMARY AND CONCLUSION

In this paper we have proved that SPS is strongly consistent in the sense that the confidence regions become smaller and smaller as the number of data points increases, and any false parameter values will eventually be excluded from the SPS confidence region, with probability one. We have also shown that a similar claim is valid for the previously proposed ellipsoidal outer-approximation algorithm. These results were illustrated by simulation studies, as well. The findings support that in addition to the attractive finite sample property, i.e., the exact confidence probability, the SPS method has also very desirable asymptotic properties.

REFERENCES

- [1] M.C. Campi, B.Cs. Csáji, S. Garatti, and E. Weyer. Certified system identification: Towards distribution-free results. In *Proceedings of the 16th IFAC Symposium on System Identification*, pages 245–255, 2012.
- [2] B.Cs. Csáji, M.C. Campi, and E. Weyer. Non-asymptotic confidence regions for the least-squares estimate. In *Proceedings of the 16th IFAC Symposium on System Identification*, pages 227–232, 2012.
- [3] B.Cs. Csáji, M.C. Campi, and E. Weyer. Sign-perturbed sums (SPS): A method for constructing exact finite-sample confidence regions for general linear systems. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 7321–7326, 2012.
- [4] M. Kieffer and E. Walter. Guaranteed characterization of exact non-asymptotic confidence regions as defined by LSCR and SPS. *Automatica*, 50:507–512, 2014.
- [5] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, 2nd edition, 1999.
- [6] Ron C. Mittelhammer. *Mathematical Statistics for Economics and Business*. Springer, 2nd edition, 2013.
- [7] Paul M.J. Van Den Hof, Peter S.C. Heuberger, and József Bokor. System identification with generalized orthonormal basis functions. *Automatica*, 31:1821–1834, 1995.
- [8] E. Weyer, B.Cs. Csáji, and M.C. Campi. Guaranteed non-asymptotic confidence ellipsoids for FIR systems. In *Proceedings of the 52nd IEEE Conference on Decision and Control*, pages 7162–7167, 2013.

APPENDIX I

PROOF OF THEOREM 2: STRONG CONSISTENCY

We will prove that for any fixed (constant) $\theta' \neq \theta^*$, $\|S_0(\theta')\|^2 \xrightarrow{\text{a.s.}} (\theta^* - \theta')^T R (\theta^* - \theta')$, which is larger than zero (using the strict positive definiteness of R , i.e., A3), while for $i \neq 0$, $\|S_i(\theta')\|^2 \xrightarrow{\text{a.s.}} 0$, as $n \rightarrow \infty$. This implies that, in the limit, $\|S_0(\theta')\|^2$ will be the very last element in the ordering, and therefore θ' will be (almost surely) excluded from the confidence region as $n \rightarrow \infty$.

Using the notation $\tilde{\theta} \triangleq \theta^* - \theta'$, $S_0(\theta')$ can be written as

$$\begin{aligned} S_0(\theta') &= R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t (Y_t - \varphi_t^T \theta') = \\ &= R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t \varphi_t^T \tilde{\theta} + R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t N_t. \end{aligned}$$

The two terms will be analyzed separately.

The convergence of the first term follows immediately from our assumptions on the regressors (A3) and by observing that $(\cdot)^{\frac{1}{2}}$ is a continuous matrix function. Thus,

$$R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t \varphi_t^T \tilde{\theta} = R_n^{\frac{1}{2}} \tilde{\theta} \xrightarrow{\text{a.s.}} R^{\frac{1}{2}} \tilde{\theta}, \quad \text{as } n \rightarrow \infty$$

The convergence of the second term follows from the component-wise application of the strong law of large numbers. First, note that $\{R_n^{-\frac{1}{2}}\}$ is a convergent sequence, hence it is enough to prove that the other part of the product converges to zero (a.s.). The Kolmogorov's condition holds since by using the Cauchy-Schwarz inequality and A4, A5,

$$\begin{aligned} \sum_{t=1}^{\infty} \frac{\mathbb{E}[\varphi_{t,k}^2 N_t^2]}{t^2} &\leq \sum_{t=1}^{\infty} \frac{\|\varphi_t\|^2}{t} \frac{\mathbb{E}[N_t^2]}{t} \leq \\ &\leq \sqrt{\sum_{t=1}^{\infty} \frac{\|\varphi_t\|^4}{t^2}} \sqrt{\sum_{t=1}^{\infty} \frac{\mathbb{E}[N_t^2]^2}{t^2}} < \infty \end{aligned}$$

Consequently, from Kolmogorov's strong law of large numbers (SLLN) for independent variables, we have

$$R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t N_t \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \rightarrow \infty$$

Combining the two results, we get that

$$\|S_0(\theta')\|^2 \xrightarrow{\text{a.s.}} (\theta^* - \theta')^T R(\theta^* - \theta') = \tilde{\theta}^T R \tilde{\theta} > 0.$$

Now, we investigate the asymptotic behavior of $S_i(\theta')$,

$$\begin{aligned} S_i(\theta') &= R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \varphi_t \alpha_{i,t} (Y_t - \varphi_t^T \theta') = \\ &= R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t \varphi_t^T \tilde{\theta} + R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t N_t. \end{aligned}$$

We will again analyze the asymptotic behavior of the two terms separately. The convergence of the second term follows immediately from our previous argument, since the variance of $\alpha_{i,t} \varphi_t N_t$ is the same as the variance of $\varphi_t N_t$. Thus,

$$R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t N_t \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \rightarrow \infty,$$

For the first term, since $\{R_n^{-\frac{1}{2}}\}$ is convergent and $\tilde{\theta}$ is constant, it is sufficient to show the (a.s.) convergence of $\frac{1}{n} \sum_{t=1}^n \alpha_{i,t} [\varphi_t \varphi_t^T]_{j,k}$ to 0 for each j and k . From A4,

$$\sum_{t=1}^{\infty} \frac{\mathbb{E}[\alpha_{i,t}^2 [\varphi_t \varphi_t^T]_{j,k}^2]}{t^2} = \sum_{t=1}^{\infty} \frac{\varphi_{t,j}^2 \varphi_{t,k}^2}{t^2} \leq \sum_{t=1}^{\infty} \frac{\|\varphi_t\|^4}{t^2} < \infty$$

Therefore, the Kolmogorov's condition holds, and

$$R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t \varphi_t^T \tilde{\theta} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \rightarrow \infty$$

Now, we show that $\|S_0(\theta')\|^2 \xrightarrow{\text{a.s.}} (\theta^* - \theta')^T R(\theta^* - \theta')$ and $\|S_i(\theta')\|^2 \xrightarrow{\text{a.s.}} 0$, $i \neq 0$, implies that eventually the confidence region will (a.s.) be contained in a ball of radius ε around the true parameter, θ^* , for any $\varepsilon > 0$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote the underlying probability space, where Ω is the sample space, \mathcal{F} is the σ -algebra of events, and \mathbb{P} is the probability measure. Then, there is an event $F_0 \in \mathcal{F}$, such that $\mathbb{P}(F_0) = 1$ and for all $\omega \in F_0$, for each i , including $i = 0$, the functions $\|S_i(\theta')\|^2$ converges.

Introduce the following notations:

$$\begin{aligned} \Gamma_{i,n} &\triangleq \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t \varphi_t^T, \\ \gamma_{i,n} &\triangleq \frac{1}{n} \sum_{t=1}^n \alpha_{i,t} \varphi_t N_t, \\ \psi_n &\triangleq \frac{1}{n} \sum_{t=1}^n \varphi_t N_t. \end{aligned}$$

Fix an $\omega \in F_0$. For each $\delta > 0$, there is an $N(\omega) > 0$, such that for $n \geq N(\omega)$ (for all $i \neq 0$),

$$\begin{aligned} \|R_n^{\frac{1}{2}} - R^{\frac{1}{2}}\| &\leq \delta, \quad \|R_n^{-\frac{1}{2}} \psi_n(\omega)\| \leq \delta, \\ \|R_n^{-\frac{1}{2}} \Gamma_{i,n}(\omega)\| &\leq \delta, \quad \|R_n^{-\frac{1}{2}} \gamma_{i,n}(\omega)\| \leq \delta, \end{aligned}$$

by using the earlier results, where $\|\cdot\|$ denotes the spectral norm (if its argument is a matrix), i.e., the matrix norm induced by the Euclidean vector norm.

Assume that $n \geq N(\omega)$, then

$$\begin{aligned} \|S_0(\theta')(\omega)\| &= \|R_n^{\frac{1}{2}} \tilde{\theta} + R_n^{-\frac{1}{2}} \psi_n(\omega)\| = \\ &= \|(R_n^{\frac{1}{2}} - R^{\frac{1}{2}}) \tilde{\theta} + R^{\frac{1}{2}} \tilde{\theta} + R_n^{-\frac{1}{2}} \psi_n(\omega)\| \geq \\ &\quad \lambda_{\min}(R^{\frac{1}{2}}) \|\tilde{\theta}\| - \delta \|\tilde{\theta}\| - \delta, \end{aligned}$$

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue. On the other hand, we also have

$$\begin{aligned} \|S_i(\theta')(\omega)\| &= \|R_n^{-\frac{1}{2}} \Gamma_{i,n}(\omega) \tilde{\theta} + R_n^{-\frac{1}{2}} \gamma_{i,n}(\omega)\| \leq \\ &\leq \|R_n^{-\frac{1}{2}} \Gamma_{i,n}(\omega)\| \|\tilde{\theta}\| + \|R_n^{-\frac{1}{2}} \gamma_{i,n}(\omega)\| \leq \delta \|\tilde{\theta}\| + \delta \end{aligned}$$

We have $\|S_i(\theta')(\omega)\| < \|S_0(\theta')(\omega)\|$ for all θ' that satisfy

$$\delta \|\tilde{\theta}\| + \delta < \lambda_{\min}(R^{\frac{1}{2}}) \|\tilde{\theta}\| - \delta \|\tilde{\theta}\| - \delta,$$

which after rearrangement reads

$$\kappa_0(\delta) \triangleq \frac{2\delta}{\lambda_{\min}(R^{\frac{1}{2}}) - 2\delta} < \|\tilde{\theta}\|,$$

therefore, those θ' vectors for which $\kappa_0(\delta) < \|\theta^* - \theta'\|$ are not included in the confidence region $\Theta_n(\omega)$, for $n \geq N(\omega)$. Finally, setting $\delta := (\varepsilon \lambda_{\min}(R^{\frac{1}{2}})) / (2 + 2\varepsilon)$ proves the statement of the theorem for a given $\varepsilon > 0$. \square

APPENDIX II

PROOF SKETCH OF COROLLARY 3

It is enough to show that $\forall i \in \{1, \dots, m-1\}$: $\gamma_i^* \xrightarrow{\text{a.s.}} 0$, as $n \rightarrow \infty$, where $\gamma_i^* = \max_{\theta: \|S_0(\theta)\|^2 \leq \|S_i(\theta)\|^2} \|S_0(\theta)\|^2$, since this implies that $r_n \xrightarrow{\text{a.s.}} 0$, where $\{r_n\}$ are the radii.

It was shown above that $\|S_0(\theta')\|^2 > \|S_i(\theta')\|^2$ (a.s.), for sufficiently large n and any $\theta' \neq \theta^*$. Then, we can also show that for $\forall \varepsilon > 0$, γ_i^* is eventually (a.s.) bounded by $\sup_{\theta: \|\theta - \theta^*\| < \varepsilon} \|S_0(\theta)\|^2$ which eventually will be bounded by $\sup_{\theta: \|\theta - \hat{\theta}_n\| < 2\varepsilon} \|S_0(\theta)\|^2$ since $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta^*$. This bound tends to zero, as $\varepsilon \rightarrow 0$, since $\|S_0(\hat{\theta}_n)\|^2 = 0$, for all n , and $\|S_0(\theta)\|^2$ is continuous w.r.t. θ . Thus, we have $\gamma_i^* \xrightarrow{\text{a.s.}} 0$. \square